

Consciousness is underived intentionality*

David Bourget

Institute of Philosophy
University of London

Noûs 44 (1):32-58. 2010.

Abstract

Representationalists argue that phenomenal states are intentional states of a special kind. This paper offers an account of the kind of intentional state phenomenal states are: they are underived intentional states. This account of phenomenal states is equivalent to two theses: first, all possible phenomenal states are underived intentional states; second, all possible underived intentional states are phenomenal states. I clarify these claims and argue for each of them. I also address objections which touch on a range of topics, including meaning holism, concept empiricism, and the relation between introspection and phenomenal consciousness. I conclude with a brief discussion of the consequences of the proposed theory for the project of naturalizing consciousness.

1. Introduction

Philosophers commonly distinguish between two kinds of mental state: intentional states and phenomenal states. The former, such as beliefs and desires, are characterized by their intentional directedness. Intentional directedness is the property of being directed at objects, states of affairs, properties, or other entities independently of the existence or instantiation of these entities. Phenomenal states, or experiences, are individuated by what it is like to be in them. The paradigmatic examples of phenomenal states are perceptual experiences, bodily sensations, and emotional feelings.

Phenomenal states and intentional states might be related in many ways. According to one view, representationalism, there is some property *F* such that the phenomenal states are one and the same as the intentional states with this property. Variants on this view have been defended by Byrne (2001), Dretske (1995, 2003), Chalmers (2004, 2006), Jackson (2004), Lycan (1996) and Tye (1995, 2000), among others. In this paper I will ask what property *F* might be.

Perhaps the leading account of property *F* is Tye's PANIC theory, which he shares in outline with Dretske.¹ According to this theory, *F* is the property of being Poised, Abstract, and Nonconceptual. Poised states are such that “they arise at the interface of the nonconceptual and conceptual domains, and they stand ready and available to make a direct impact on beliefs and/or desires” (Tye 2000, p. 62). On Tye's usage, intentional states are abstract when their contents are not object-involving. Nonconceptual states are states one can be in without possessing concepts for the entities that figure in their contents.

The PANIC theory faces some difficulties. For example, one might reasonably doubt that consciousness involves something as vague as poisedness. Thought experiments of the kind

discussed by Block (1995) also suggest that it is possible to have experiences which are fully disconnected from one's higher cognitive faculties, contrary to the requirement of poisedness.²

My aim in this paper is to propose an alternative account of the distinguishing property F. All representationalists should agree that part of what it takes for an intentional state to be conscious is that its intentionality be original or underived. I suggest that we need not add anything else to our account of consciousness: the phenomenal states are simply the underived intentional states. This proposal captures the spirit of the view Chalmers (2004) calls “pure representationalism”, in that it appeals to a minimum of non-representational elements to distinguish phenomenal and non-phenomenal intentional states. As a reminder of this, I refer to it as *the PURE theory of experience*. The acronym “PURE” stands for *Pure Underived Representational Events*, which is what the theory says experiential events are (“pure” means “nothing more than”).

I will begin by explaining the PURE theory in more detail. Then I will give an argument for it before turning to the main objections it faces. I will not defend representationalism *per se*, but I will argue that once this view is accepted, the property of being underived can play the role of property F.

2. The PURE theory

The PURE theory says that the phenomenal states are one and the same as the underived intentional states. It is equivalent to the conjunction of two theses. First, the *originality thesis*:

The originality thesis: all possible phenomenal states are underived intentional states.

The second thesis is the converse, namely, that all possible underived intentional states are phenomenal states. We can also put it like this:

The derivation thesis: all possible non-phenomenal intentional states are derived.

I must now clarify what I mean by “derived” and “underived”. The difference between derived and underived representations is normally glossed as that between those representations which are intrinsically representational (the underived ones) and those which are not (the derived ones; c.f. Searle 1983 and Strawson 1994). I take this to mean that what distinguishes derived from underived representations is that the former, but not the latter, have their contents or representational properties at least in part in virtue of intentional states distinct from themselves, or relations to such states. For example, linguistic representations seem to differ from underived mental states in that they have content at least in part in virtue of their users' intentional states or relations to their users' intentional states. This is the understanding of the derived / underived distinction I will adopt for the purposes of this paper: a representation (either a state or an object) is derived just in case it is such that it must be realized at least in part by intentional states distinct from itself, or relations to such states.³ Underived representations are representations which are not derived. According to the PURE theory, a phenomenal state is an underived intentional state, that is, an intentional state which need not be realized by any other intentional states or relations to other intentional states.

The claim that all non-phenomenal intentional states are derived (the derivation thesis) might seem surprising given that the paradigmatic examples of derived representations are linguistic. But the derived / underived distinction leaves room for derived mental states. I will now survey four different kinds of derived intentionality we might find among mental states. This will give more shape to the PURE theory and help make sense of the derivation thesis. At this stage, however, I am not trying to argue for the latter thesis, but only to give the reader a sense of the various ways the PURE theory could be developed, some of which I will pursue later in the paper.

(1) *Deferential derivation.* Some mental states might get their contents from the mental states of members of linguistic communities in the same kind of way as linguistic representations. Take a child who is learning to speak. If she said, sincerely, “the barrister wasn't nice to daddy”, she could plausibly be described as believing that the barrister was not nice to her father. This would be so even if she had a very imperfect command of the meaning of “barrister”. In such a case it would seem that the contents of her speech act and the thought she expressed derived from the linguistic conventions by which she was bound. They would therefore seem to be ultimately derived from others' mental states. It may be that many of the contents we express and believe derive in part from others' mental states in this way. This appears to be the upshot of Burge's (1979) anti-individualism and Putnam's (1975) argument for the division of linguistic labor.

(2) *Compositional derivation.* Perhaps the most easily recognized kind of derived intentionality is that of composite states. Take the state of experiencing-redness-while-believing-that-P (for some P). This is arguably an intentional state: it has intentional directedness at least toward P. But it must be realized by a belief and an experience in order to be instantiated. This means that it is derived. It is derived in virtue of having other intentional states as components.⁴

(3) *Dispositional derivation.* Another way some intentional states might derive from others is through causal connections. The best examples here are non-occurrent states—states which involve no mental “goings-on”. Arguably, almost all belief ascriptions are ascriptions of non-occurrent states. For example, it could be true of me at a given time that I believe monkeys like bananas even though I am deep asleep and have nothing “on my mind”. Moreover, non-occurrent states seem to be constituted at least in part by dispositions to have relevant occurrent states. We would not say that I believe monkeys like bananas if I were not disposed to

think (occurrently, with a certain conviction) that monkeys like bananas upon being suitably prompted in the right conditions. It is on the face of it at least in part in virtue of my disposition to token relevant occurrent states that I count (while asleep) as believing that monkeys like bananas. If this observation extends to all non-occurrent states, then non-occurrent states merely have derived intentionality.

(4) *Matching derivation*. A fourth, potentially more controversial kind of derived intentionality is best brought out by means of the distinction between wide and narrow content drawn by Chalmers (2002a), Jackson (2003), Loar (1987), Segal (2000), and many others. To illustrate the distinction, take a typical token s of the sentence “Water is H_2O ”. According to the preceding theorists, s has two contents, one wide and one narrow. Its wide content is the trivial proposition H_2O is H_2O , while its narrow content for an average individual might be a descriptive proposition of the form *the actual watery stuff is the stuff scientists call “ H_2O ” in this world*, a relation between clusters of properties, or some such. The wide content of s , which depends on factors external to the speaker's brain, accounts for the fact that it is necessarily true. By contrast, its narrow content is determined by the state of the speaker's brain and might account for the fact that it appears cognitively significant or *a posteriori*. The details of this view vary from one author to the next, but a common thread is that the narrow content of a representation at least partly determines its wide content through something like descriptive matching. For example, one might say that the term “water” has H_2O as wide content at least in part in virtue of the fact that the latter matches the properties figuring in its narrow content. The same could be true of mental states: perhaps thoughts with H_2O content have it in part in virtue of other thoughts having watery stuff content. The former would be *matching* states and the latter *matched* states; the former would be derived, while the latter could be either derived or underived.⁵

Let us take stock. There are many ways intentional states could derive from one another (realize one another), but four possible derivation mechanisms are salient: derivation through deference, composition, dispositions, and matching. I suggest that the phenomenal states are precisely the underived intentional states, that is, the intentional states which need not be derived in any of these ways, or any other.

3. Argument

In this section I argue for the originality and derivation theses I introduced at the beginning of the preceding section. These theses together entail that, necessarily, all and only phenomenal states are underived intentional states (the PURE theory).

3.1. The originality thesis: all possible phenomenal states are underived intentional states

The originality thesis can be factored into two claims: first, that all possible phenomenal states are intentional states; second that all possible phenomenal states are underived if intentional. I will not argue for the first claim here. It is a central tenet of representationalism. As I said in introduction, I take this view as background assumption for present purposes. I will focus on the claim that phenomenal states are underived.

It seems plausible that any sensory experience could in principle occur in isolation from all other intentional states. Take for example my current experience of the whiteness of my computer's screen (call it *e*). On the face of it, a being could pop into existence, have a token of *e* for a split second, then disappear without having had any other intentional states or stood in any nontrivial relations to other intentional states. The cogency of this scenario is at least good evidence of its metaphysical possibility. It therefore seems that a token of *e* could occur without being realized by distinct states or relations to distinct states. Now, an underived state is a state whose tokens need not be realized by distinct states or relations to such states. It fol-

lows that *e* is underived. Since this reasoning straightforwardly extends to other sensory experiences (experiences of sounds, shapes, smells, etc), it seems probable that sensory experiences are underived.

This reasoning also appears to carry over to non-sensory experiences, e.g., the tip-of-the-tongue feeling, the feeling of confusion, the experience of familiarity, the experience of an affordance, and the experience of causation. Take causation. Assuming one can experience causal relations, it seems possible to be phenomenally conscious merely as of *something causing something* without undergoing other experiences or having any other intentional states. One might describe such an experience as a vague feeling that something is happening, that a force was exerted, or some such. It therefore seems that any phenomenal state could occur in isolation from any other intentional states, hence that all possible phenomenal states are underived.

In case one is suspicious of *a priori* assessments of metaphysical possibility, there is a largely empirical line of argument to the same conclusion. It is well known that human brains are far more similar to the brains of other species at the level of perceptual systems than at the level of cognitive systems. Consequently, it is reasonable to expect there to be creatures capable of sensory experiences similar to ours even though they have almost none of the cognitive capabilities we have (perhaps some birds, some fishes, or some dim mammals are like that). This would strongly suggest that sensory experiences can occur without thoughts. The fact that sensory experience is largely independent of influence from background beliefs (as cases of familiar perceptual illusions show) likewise suggests that sensory experience is independent of thought.

There are also empirical reasons to think that sensory experiences are independent from one another. For a start, our sensory modalities are largely modular: some people can see but

not hear, others can smell but not see, touch but not smell, and so on. Pathological cases further show that there is a great deal of modularity *within* sensory modalities. For example, achromats can see shapes but not colors. Cases of visual agnosia provide more striking examples of how aspects of visual experience can come apart. For instance, some individuals who suffer from visual form agnosia can recognize movement but not discriminate shapes (Farah 1990, p. 12). Others who are afflicted with dorsal simultanagnosia can see certain objects' individual features without being able to grasp them as wholes, and some lack the ability to visually recognize the presence of more than one object at a time (to experience numerical distinctness; Farah 1990, pp. 27-8). Farah discusses many other kinds of agnosia. These cases, along with functional imaging experiments and single-cell recordings, show that vision is highly modularized (Cooney & Gazzaniga 2003). Of course, they do not demonstrate the metaphysical possibility of any sensory experience occurring independently of all others; no amount of empirical data can show this. Nevertheless, the fact that sensory experiences are by and large nomically dissociable from one another suggests that one could in principle have any sensory experience in isolation from all other possible sensory experiences. Given that sensory experiences are also independent from thoughts, it is tempting to conclude that any sensory experience could in principle occur in isolation from all other possible intentional states. It would then follow that sensory experience is underived.

It remains to see if there is also empirical evidence that non-sensory experience is underived. To continue with the example of causation, we have at least two relevant pieces of empirical evidence regarding the relation between experiences of causation and other experiences: 1) it is in practice possible to experience causation without experiencing familiarity, affordances, or having any other non-sensory experiences; 2) the sensory experiences which accompany experiences of causation vary greatly from case to case. These two facts together suggest that experiences of causation could in principle occur in isolation from any other ex-

periences. Now, it also seems possible in practice to experience causation without having any relevant thoughts. For example, suppose that I am unintentionally kicking small stones as I am walking down a gravel path. I could in practice be dimly aware of this fact and have a certain feeling of causing change in my environment without having any potentially relevant thoughts, or even dispositions to have relevant thoughts. This suggests that I could in principle experience causation without having any thoughts at all. Given that I can in principle also experience causation without having any other experiences, it is reasonable to conclude that one could in principle experience causation without tokening any other intentional states. This would mean that experiences of causation are underived. This argument appears to generalize to other non-sensory experiences. Since we have already seen that sensory experiences are probably underived, it seems likely that all experiences, either sensory or non-sensory, are underived.

To summarize, there are two sources of evidence for the originality thesis: modal intuition and empirical data. Both suggest that any experience could in principle occur in complete isolation from all other possible intentional states.⁶ If this conclusion is correct, all phenomenal states are underived, because it is sufficient for a state to be underived that it be possible for it to occur without being realized by distinct intentional states.

A natural objection to the originality thesis stems from the view that “experience is conceptual”. This view suggests that the content of experience derives from concepts. However, it does not entail that experiences derive from other intentional states, so it is compatible with the originality thesis.

There are two different interpretations of the claim that “experience is conceptual” (see Heck 2000). On the *content view*, a state is conceptual just in case its object (its content) is an entity of a special kind, e.g., a Fregean sense. On the *state view*, a state is conceptual when it

is of necessity realized by (or dependent on⁷) representational vehicles of a special kind, e.g. a special sort of term in the language of thought. It is true on each view that conceptual states “involve concepts”. However, this claim and the term “concept” take different meanings depending on what being conceptual is supposed to amount to. On the content view, conceptual states involve concepts in that they represent concepts (contents of a special kind); on the state view, they involve concepts in that they are realized by concepts (representational vehicles of a special kind).

The content view is compatible with the originality thesis. On the face of it, at least, there is no reason to think that a state whose content involves something like a Fregean sense has to be realized by distinct states. Regarding the state view, the mere fact that experiences are realized by special representational vehicles does not conflict with the originality thesis. It would conflict only if concepts were themselves intentional states, which they are not (they are mere vehicles). Now, one might think that concepts derive their contents from relations to thoughts, i.e. from their conceptual roles. The state view combined with a conceptual role semantics for the concepts relevant to experience would conflict with the originality thesis. But a theory of content according to which the relevant concepts get their contents solely through causal relations with their objects would not have this consequence. By itself, the state view is compatible with the originality thesis. More importantly, conceptual role semantics does not seem plausible as a theory of the contents of experience (see §4.3 below).

3.2. The derivation thesis: all possible non-phenomenal intentional states are derived

The derivation thesis is a weakened version of the following claim:

The strong derivation thesis: all possible non-phenomenal intentional states derive from phenomenal states.

A number of authors have recently expressed sympathy for the strong derivation thesis (e.g., Graham, Horgan & Tienson 2007; Kriegel 2003; McGinn 1988; Searle 1990, 1992; Strawson 1994). It has also been widely held during the past few hundred years: British empiricists were explicitly committed to it; Kant famously held that concepts without intuition are blind, which can at least be read as suggesting a similar thesis; Brentano described mental states as “presentations” (experiences) and states “dependent on presentations”⁸; Russell appears to have been committed to some kind of derivation thesis given that his notion of acquaintance is akin to our notion of consciousness.⁹ These authors have offered a range of reasons for holding the strong derivation thesis, and these reasons all carry over to the weaker derivation thesis.

Here, however, I want to give an argument for the (weak) derivation thesis which does not rely on the strong derivation thesis. The argument goes as follows:

P1. Every possible underived intentional state is either an attitude or a phenomenal state.

P2. Every possible attitude is derived.

Therefore, every possible underived intentional state is a phenomenal state.

The conclusion of this argument is equivalent to the derivation thesis (all possible non-phenomenal intentional states are derived). I will begin by explaining what I mean by “attitude”. Then I will say why I find P1 and P2 plausible.

All attitudes are directed at entities of some kind or other. Perhaps the most commonly discussed objects of attitudes are propositions. An attitude toward a proposition is what we call a *propositional attitude*. The paradigmatic examples of propositional attitudes are belief and desire, but propositional attitudes are far more diverse than this. Examples of propositional attitudes include: wondering whether P, rejecting the claim that P, suspecting that P, entertaining

the possibility that P, being intrigued by the fact that P, assuming that P, wishing that P, being frustrated by the fact that P, being ashamed that P, being anxious that P, being afraid that P, being thrilled that P, intending to make it so that P.

Other attitudes are directed at objects or properties rather than propositions. Examples of such *sub-propositional attitudes* include believing someone, being inquisitive about someone, being puzzled by an object, ignoring someone, liking something, loving someone, hating someone, fearing something, wanting something, admiring someone, etc. It is arguable that sub-propositional attitudes are very often derivative on propositional attitudes. For instance, I fear Rufus at least in part in virtue of the fact that I fear being bitten by him (that I will be bitten by him).

By “attitudes”, I mean states of the most specific natural kind that includes all the preceding examples. This kind of state appears to be characterized by three features. First, states of this kind are attitudes in the ordinary sense: they are attitudes in the same sense that being cheerful and being arrogant are attitudes (they are “ways of being”). Second, they have a kind of directedness (which cheerfulness and arrogance arguably lack). Third, they are intentional. This is to say that the sort of directedness they involve can obtain even if the entities they are directed toward do not exist, are not instantiated, or do not obtain (depending on what kind of entity they are). I will say more about the nature of attitudes as we progress.

The first premise of my argument for derivation (P1) is that every possible underived intentional state is either an attitude or a phenomenal state. This might seem to be true simply in virtue of the fact that every intentional state (whether derived or underived) is either an attitude or a phenomenal state. However, it is important that we allow derived states which are neither attitudes nor phenomenal states in order to account for composite states of the kind I mentioned in §2. For example, one might count the state of doubting-some-content-one-is-ex-

perceiving as an intentional state. It involves more than an attitude and more than a phenomenal state. This means that it is neither a phenomenal state nor an attitude. However, it is a derived state, because it must be realized by an attitude and a phenomenal state. The occurrence of “underived” in P1 is there to accommodate such composite states. P1 seems prima facie plausible because all non-composite intentional states appear to be either attitudes or phenomenal states.

In order to see better the plausibility of P1, try to picture Void, a being capable of intentional states but constitutively incapable of having experiences or attitudes. Void cannot have any beliefs, desires, wishes, fears, pains, feelings, and so forth, *but it has mental states*.¹⁰ I find myself unable to imagine Void as described. As far as I can tell, there is no Void in logical space. This is not surprising, because even when we ascribe mental states to machines or other simple systems, we ascribe them attitudes and phenomenal states. For example, we say that computers believe, learn, and see; that immune systems learn, sense, and want; that flowers sense and know things. If personal computers, immune systems, and flowers could not be said to have attitudes or experiences—not even metaphorically—we would not even be tempted to ascribe them mental states as a *façon de parler*. Trying to imagine a being which has mental states (let alone intentional states) but lacks the capacity to have attitudes or experiences is like trying to imagine someone walking without causing anything to move. In absence of evidence to the contrary, this is good reason to suppose that all possible mental states either are, or are partly constituted by, attitudes or experiences. This in turn would mean that every underived intentional state is either an attitude or a phenomenal state.

One might ask why *thinking* and *entertaining a proposition* are not counterexamples to P1. “Think” can take two different meanings depending on how it is used. When we say that someone *thinks that P*, we mean that she believes P to be true. For example, I think that we

are headed toward a recession. This is to say that I believe we are headed toward a recession. Thinking is in this sense an attitude, because believing is an attitude. We can use “think” in a different sense, as when we say that someone is *thinking about the fact that P*. Here “think” is used to mean that one is engaged in some kind of mental process involving representations of certain contents. One may or may not count the process of thinking as a mental state. If it is counted as a (temporally structured) state, it is a derived one, because a reasoning process is of necessity realized by the states it connects. If thinking in the present sense is not a state, then it falls outside the scope of P1, which bears only on intentional *states*. Either way, no case of thinking about a content contradicts P1. “Entertain” means something similar to “think about”. When we say that someone is entertaining a claim or possibility, we mean that she is engaged in some kind of exploration of its plausibility or relevance. Entertaining a proposition is a process which could be classified either as not a state or as a temporally structured, derived state.

I now turn to the claim that all possible attitudes are derived (P2). At this stage we need to ask in virtue of what we have beliefs, desires, and other attitudes. This is of course a difficult question. Fortunately, we need not provide a full answer to it; we need not commit on necessary and sufficient conditions for attitudes. It is enough for our purposes to ask whether *part* of what it is for a state to be a belief, desire, or attitude of some other kind is for it to stand in certain relations to other intentional states (either attitudes or phenomenal states). We might call the view according to which this is so *minimal functionalism*. If minimal functionalism is correct, attitudes are realized at least partly by relations to distinct intentional states, so are derived.

Perhaps the best way to bring out the plausibility of minimal functionalism is to compare attitudes with phenomenal states. Phenomenal states are intentional states, but one can experi-

ence a given content without also having a belief, desire, or any other attitude toward it. Imagine for example that you are looking at a stick which is half immersed in water. Even if your experience presented the stick as bent to you, you would not necessarily believe, wish, suppose, desire, or intend it to be bent, or have any other attitude toward its being bent. Alternatively, suppose that you are lying in bed with your eyes closed and a mental image of a beach pops in your mind. It does not follow from the fact that you have a phenomenal representation of a beach that you have any attitude whatsoever with respect to any beach. Experiences are intentional states, but they are not attitudes in the sense I am concerned with (see §4.1 below for more on this point).

What is missing from experiences to confer attitudes? In particular, what else would it take for you to count as believing that the stick is bent, or as desiring to be at the beach? Dispositions to behave (move) in certain ways would seem insufficient. For not believing that P seems entirely consistent with experiencing P and having exactly the behavioral dispositions of someone who believes that P. It is a familiar point that the very same behavioral dispositions can be explained just as well by mutually exclusive sets of attitudes, e.g., a set involving the belief that P and other attitudes, and a set involving the belief that not-P among other attitudes. This is so quite independently of whether one represents P consciously or not. But if behavioral dispositions would not be sufficient for you to believe that the stick is bent or desire being at the beach, what is missing has to be other mental states or dispositions to token other mental states. Perhaps you would have to feel or be disposed to feel certain things—a certain conviction for belief, a certain pull for desire. Or perhaps what you would need is dispositions to token attitudes consistent with believing that the stick is bent or desiring to be at the beach. Or one might think that either dispositions to have relevant feelings or dispositions to token relevant attitudes would do it.

Whatever the case may be, the preceding examples bring out the fact that believing and desiring are attitudes in the ordinary sense of “attitude”. Your beliefs and desires are not just static representations in your brain. They also involve (if they are not exhausted by) complex dispositions to token certain patterns of mental states (phenomenal states and other attitudes). As such, attitudes are parasitic on the states they connect. Hence minimal functionalism and the derived status of attitudes.

It is worth emphasizing how weak minimal functionalism is. Since it only implies that certain relations to distinct intentional states are *partly* constitutive of attitudes, it is not open to objections to functionalism which purport to show that functional role is *insufficient* for attitudes. For example, it is not open to Searle's (1980) Chinese room and Block's (1978) China brain objections. It is also not open to Galen Strawson's (1994) Weather Watcher argument against “neobehaviorism”, because it does not draw any necessary connection between mental states and behavior. In fact, Strawson appears to endorse a form of minimal functionalism, at least as far as the attitude of desire goes:

... the primary linkage of the notion of desire to a notion other than itself is not to the notion of action or behaviour but rather to the notion of being pleased or happy or contented should something come about ... (p. 280)

Strawson implies that it is part of a state's being one of desire that its owner be disposed to token a certain kind of phenomenal state or attitude were it to be satisfied. This is a variant on minimal functionalism about desire.

Strawson is not the only one with sympathies for minimal functionalism. It is part of the full-blown functionalist view of attitudes, whose proponents are many, e.g., Armstrong (1973), Audi (1972), Baker (1995), Fodor (1990), Harman (1973), Lycan (1982), Pettit (1993), and Stalnaker (1984), to name a few. Dennett's (1987) and Davidson's (1984) brands

of interpretationism also involve a minimal functionalist element. Likewise, the teleosemantic theory defended by Millikan (1984) entails that attitudes are constituted in part by relations to distinct intentional states (they must form systems of “intentional icons”). As far as I can tell, almost all mainstream theories of propositional attitudes embody a form of minimal functionalism.

One might question my claim that functionalism (either full or minimal) entails that attitudes are derived. For take a standard wide functionalist view according to which attitudes are states that stand in certain causal relations to other mental states and external entities. One way to state such a view about belief would be to say that to have a belief is to be in a state *s* such that there are other states and entities *t*, *u*, *v*, ... such that *s* stands in a complex causal relation *R* to the latter. The right hand side of this identification is couched in non-mental vocabulary. It might therefore seem that what realizes beliefs on the functionalist view are mere patterns of neutrally characterizable states, not relations to other mental states. If this were so, functionalism would not entail that attitudes are derived. Since minimal functionalism is strictly weaker than functionalism, this would show also that the former does not entail that attitudes are derived.

I agree that belief can be identified with a neutrally characterizable state on the standard functionalist view, but it remains that this state is constituted by relations to states which are, in light of the same view, distinct intentional states. It matters not what kind of vocabulary the relevant states can be described in. It is enough for a state to be derived that it be of necessity realized by relations to distinct intentional states, so attitudes are derived on the functionalist view. Note also that the objection proves too much: if being of necessity realized by relations to states of a certain type *S* were insufficient to be derived when the *S* states are reducible to pure functional states, then functionalism would have the consequence that linguistic repres-

entations are underived.

This concludes my argument for the derivation thesis and the PURE theory. The overall argument is relatively straightforward. First, our rational ability to assess metaphysical possibility *a priori*, combined with relevant empirical facts, provides significant evidence that phenomenal states are underived (the originality thesis). Second, intentional states appear to fall into three categories: phenomenal states, attitudes, and composite (derived) states. Minimal functionalism, an intrinsically plausible and widely held view, tells us that attitudes are parasitic on other states (that they are derived). It follows that non-phenomenal states are derived (the derivation thesis). We therefore have at our disposal a simple yet largely reductive account of what distinguishes phenomenal states from other intentional states: all and only phenomenal states are underived. We need not appeal to unwieldy notions such as poisedness and nonconceptualness in order to explain consciousness, because consciousness is simply underived intentionality.

4. Objections

4.1 Phenomenal attitudes

One might think that some attitudes are also phenomenal states. Let us call an attitude which is also a phenomenal state a *phenomenal attitude*. If there were phenomenal attitudes, it would follow from the originality thesis that they are underived. But I claimed as part of my argument for the derivation thesis that all attitudes are derived (P2). Either there are no phenomenal attitudes, the originality thesis is false, or my argument for derivation is unsound.¹¹ Let us look into the main reasons why one might think there are phenomenal attitudes.

Representationalism is sometimes described as the view that phenomenal states are a special class of attitudes (c.f., Crane 2003, Jackson 2004). On this reading, representationalism

commits one to phenomenal attitudes. But the representationalist view stated in introduction only entails that phenomenal states are intentional states. Whether it follows from this that phenomenal states are attitudes depends on whether all intentional states are attitudes. Here we run into terminological difficulties. It seems to me that those who identify phenomenal states with a special class of “attitudes” are using the term “attitude” in a sense different from mine. Take Crane (2003), for example. On his usage, a propositional attitude is a state of being related through an “intentional mode” to a content. Crane does not say explicitly what intentional modes are, but he counts sensing painfully as an intentional mode. It seems that an intentional mode is nothing more than a mental relation involving intentional directedness. Attitudes would in this sense encompass all intentional states in my sense. But they are not obviously the same as attitudes in my sense. Remember that an attitude in my sense is a state of the most specific natural kind that encompasses all the examples I gave in §3.1, which include no experiences. It is plausible that attitudes in this sense constitute a narrower category than what I call “intentional states” and Crane refers to as “attitudes”. This possibility is at least not ruled out *a priori*. As a result, it does not seem that representationalism commits us to phenomenal attitudes (in my sense of “attitude”).

Now, one might think that there are phenomenal attitudes even if representationalism does not commit us to their existence. Arguably the best potential examples of phenomenal states which are also attitudes are felt urges. Suppose for instance that I felt a strong urge to eat a piece of black forest cake someone put in front of me. This feeling would be a phenomenal state. It might also seem to be a kind of attitude—a desire-like attitude. Felt urges might therefore seem to be phenomenal attitudes. Similarly, one might think that a perceptual experience accompanied by a certain feeling of conviction ought to count as a belief-like phenomenal attitude.

Against this, imagine that I felt an urge to eat the cake but had no corresponding cognitive or behavioral dispositions: though I felt an urge to eat the cake, I gave strong nonverbal signs of loathing it, asked the person next to me if she would not want my share, and was profoundly relieved I would not have to suffer its sweetness when she accepted it. The natural thing to say in this case would be that I momentarily felt *as if* I wanted the cake but did not really want it. The same is true of belief-like attitudes and the feelings which sometimes accompany them. For instance, a claim can momentarily feel right to you without your believing it. One can momentarily feel as if one believed something without really believing it. This suggests that feelings are not attitudes, but merely normal concomitants of attitudes.

It is worth noting also that the present difficulties can be avoided without denying the existence of phenomenal attitudes. We need simply change P2 for the claim that *non-phenomenal* attitudes are derived. This would make it compatible with the originality thesis and the existence of phenomenal attitudes. Moreover, the derivation thesis would still follow from this weakened premise and P1: if every underived intentional state is either an attitude or a phenomenal state and every non-phenomenal attitude is derived, every underived intentional state is a phenomenal state. I don't think any phenomenal state is properly counted as an attitude, but one could allow this at little cost to the PURE theory.

4.2 Emotions

Emotional feelings such as the feelings of joy and fear are paradigmatic phenomenal states. So the PURE theory implies that they are underived. But (as an anonymous referee notes) it is a common view that the intentionality of emotions derives from that of related beliefs, desires, or other propositional attitudes. This kind of cognitivist view might seem to conflict with the originality thesis for emotional feelings.

Here it is important to distinguish between emotional feelings and other, more complex emotional states. I reserve the term “emotional feeling” for the *phenomenal states* associated with emotions. Most authors agree that emotional states generally go beyond emotional feelings (see de Sousa, 2007). Perhaps the best examples of emotional states which involve more than emotional feelings (more than phenomenal experiences) are love and hate, two paradigmatic emotional states. These states, or their full-blown variants at least, have aspects which cannot be reduced to pure phenomenal experience. For instance, one might reasonably think that certain cognitive dispositions are essential to loving or hating someone. I will call “complex emotions” the emotional states such as love and hate which go beyond emotional feelings. I suspect that most states we consider emotional fall in the category of complex emotions, not emotional feelings. It is complex emotions which generally interest us most in theoretical discussions as in everyday life.

Now, the sort of cognitivist view alluded to above is not as plausible when applied to emotional feelings as it is when applied to complex emotions. It seems plausible that complex emotions derive in part from relations they stand into with distinct attitudes. On the other hand, emotional feelings seem independent of attitudes for reasons parallel to the considerations I offered in defence of the originality thesis in sections 3.1 and 4.1: it seems that a being could in principle have any emotional feeling without having any attitudes or distinct intentional states, or dispositions to token such states. It therefore seems that only complex emotions derive their intentionality from distinct attitudes. Complex emotions are not phenomenal states because they are not individuated merely by what it is like to have them, so this minor restriction to the cognitivist view makes it compatible with the PURE theory.

4.3. Meaning holism

Meaning holism may be glossed roughly as the view that mental representations necessarily

come in large packages. Given that experiences are mental representations, someone who holds this could be inclined to think that they are derived.¹² This objection to the originality thesis might seem particularly pressing in light of my appeal to minimal functionalism, which appears close to meaning holism.

My first response to this objection is that the well-known arguments for meaning holism do not even purport to be arguments for holism about the contents of phenomenal states. All arguments for meaning holism are in fact arguments for *attitude* holism. They might have been glossed as or taken to be arguments for holism about semantics, intentionality, or representation generally by Block (1993), Davidson (1984), Lewis (1974), and Searle (1983), but it is striking that none of these philosophers thinks of experiences as intrinsically representational. As far as I know, no one has put forward an argument intended to show that the contents of experience are holistic.¹³ To put the point differently, the name “meaning holism” appears to be a leftover from the days when we thought of consciousness on the model of contentless raw feels.

Besides the lack of an argument purporting to show that experiences are holistic, it seems clear that extant arguments for attitude holism do not apply to experiences. Here I will limit myself to what appears to be the two main lines of argument for holism.

The first is that of Davidson and Lewis, which takes its starting point from the view that “certain [holistic] principles of charity are constitutive of content ascription” (Fodor & Lepore 1992, p. 59). The minimal functionalism about attitudes I briefly defended above has some affinities with this claim, in that it portrays attitudes as essentially involving broadly speaking rational connections between mental states. However, experiences differ from attitudes in this regard: it does not seem that an experience's content depends on how it fits within one's rational economy over time. Take for example Flashman, a Swampman who survives only long

enough to have a few experiences. Flashman might perhaps be incapable of attitudes because he cannot be charitably interpreted as having any. But most would balk at the suggestion that it is metaphysically impossible for Flashman to have experiences. This illustrates the fact that the relevance of charity ends at the boundary between attitudes and phenomenal states.

Perhaps the main motivation for holism is conceptual role semantics, which is at least non-atomistic by design. Whether this line of argument extends to phenomenal states or not depends on whether it is plausible that the content of an experience is determined by its inferential role. But the richness of experience gives us reason to think that inferential role does not determine its contents. Take for example the total, unified visual experience I have of my office at this time (call it *o*). Millions of different shades and even more small surfaces figure in the content of *o*. Now take an experience *o'* which is just like *o* except that its content represents the top-left most perceptible point on the copy of Hume's *Treatise* two meters away from me as being of a slightly different shade of purple. I doubt that one's inferential dispositions with regard to experiences such as *o* and *o'* are always different. In the same vein, cases of change blindness with respect to striking features of objects of attention seem to show that some differences in phenomenology (even quite stark ones) sometimes fail to be reflected in the connections phenomenal states stand in with other states (see Levin et al 2002 for some relevant cases).¹⁴

Let me emphasize that I do not wish to criticize holism, Davidsonian charity, or conceptual role semantics as far as their applicability to attitudes goes. I am happy to grant that meaning holism and the views which support it apply to attitudes—they provide more evidence for minimal functionalism. However, I believe that an important difference between attitudes and experiences is reflected in the fact that these views seem applicable to the former but not to the latter.

4.4. The objection from blindness

It is tempting to suppose that derived states ultimately derive from underived states; otherwise we would have to contend with infinite or circular derivation chains. So the (weak) derivation thesis I argued for leads naturally to the strong derivation thesis I mentioned at the beginning of §3.2. It might also seem plausible that derived states must ultimately derive from underived states *with the same contents*. More specifically, one might think that, given the PURE theory, non-phenomenal states which share content with experiences should derive their contents from these experiences. However, this leads to an objection: congenitally blind people are able to think about colors like everyone else even though they have no color experiences to derive the contents of their color thoughts from; where do they get these contents from?

My view is that blind people's color thoughts are either deferential or matching states (or both). Note first that those who are incapable of experiencing colors (either in perception or in imagination) seem to be missing something in the neighborhood of cognitive representational abilities. Here for example is what Knut Nordby, a complete achromat, reports:

Although I have acquired a thorough theoretical knowledge of the physics of colours and the physiology of the colour receptor mechanisms, nothing of this can help me to understand the true nature of colours. From the history of art I have also learned about the meanings often attributed to colours and how colours have been used at different times, but this too does not give me an understanding of the essential character or quality of colours. (Nordby 1990)

Nordby claims to be lacking a full grasp of the nature of colors. I am inclined to believe him because when I want to form as good an idea as I can of a color, I look for a sample or engage in imagery. Since he can neither see nor visualize colors, Nordby is unable to use the method which seems to give me the best grasp I can have of them. He must be missing some-

thing.

To see more precisely what it is that an achromat or blind person is incapable of, consider the following scenario. Bob has been blind from birth and so has never had the ability to experience color (not even in imagination). One day, a cosmic accident occurs that has two unfortunate consequences. The first consequence is that the colors of objects become permanently inverted all over the world. The second is that Bob is knocked into a deep sleep for several weeks. During this time, an international convention takes place at which it is decided that the meanings of color terms would henceforth be inverted (so that one can again say truly that ripe bananas are “yellow”). It takes some adjustment, but by the time Bob wakes up everyone has adapted to the new linguistic conventions. Several years down the road, but before he has realized that naming conventions for colors have changed, Bob gets into an argument with Alice about the color of the walls in his mother's dining room. He thinks he remembers recently hearing the walls described as “pink”. As a result, he expresses his view on the matter as follows:

Bob: The walls are pink.

Alice, who is normally sighted, thinks she has a visual memory of the walls being the color she describes as “pale green”.

Alice: No, they are pale green.

As it happens, the walls used to be pale green but are now pink due to the world-wide color inversion. Since the color naming conventions Alice is bound by have also been inverted, she is right to say that the walls are “pale green”. This seems to be a situation in which she is right *and Bob is wrong*.

Had naming conventions not been changed or colors not been inverted (but not both), Bob

would have been right. This suggests that his color thoughts have their contents in part in virtue of deference or descriptive matching. For example, his belief that the color green is common in nature might be derived from a distinct thought with a content of the form *the color actually called “green” is common in nature* or *the actual color of cucumbers is common in nature* (these examples are of course simplistic). Alternatively, it could get its content directly from linguistic conventions and sighted speakers' color thoughts.

That Bob's color thoughts are derived in one of these ways can be made clearer by contrasting his case with that of Chris. Chris is normally sighted. Like Bob, he is knocked into a deep sleep by the color-inverting cosmic accident and misses out on the ensuing terminological reform. To continue the parallel, let us say that Chris fails to notice that colors and color naming conventions have been inverted during his absence because he is participating in a long-term experiment requiring him to wear contact lenses which make him see everything in black and white. He, too, has recently heard the walls of a certain dining room described as “pink” and engages in an argument with Alice on this matter. Chris says that the walls are “pink”, Alice says that they are “pale green”. Alice does not know that Chris is unaware of the new color naming conventions. As a result, she is not trying to adapt the vocabulary she uses when talking to him.

In this case we would expect that Chris and Alice are not really disagreeing but merely using the terms “pink” and “pale green” differently—we would expect them to be both thinking of the walls as pink. This is especially clear given that we would expect them to be both entertaining, or ready to entertain, mental images of a pink dining room. Think also about how you would intervene in this dispute. Knowing that Chris is normally sighted but unaware of the new naming conventions, you would suggest that he and Alice clarify their terminologies first.

A natural explanation of the difference between Bob and Chris is that Chris, but not Bob,

has no need for descriptive or deferential referential mechanisms—he grasps the natures of colors directly through his ability to token phenomenal representations of them. By contrast, Bob merely grasps that they are called certain names and related to objects in certain ways. Bob's color thoughts appear to be derived, but not from any states of his with the same contents.

4.5. Concept empiricism

I mentioned earlier that the weak derivation thesis could reasonably be taken to imply the strong derivation thesis (that all non-phenomenal states derive from experiences). This thesis might in turn seem to suggest a kind of Humean concept empiricism. To a first approximation, Humean concept empiricism is the view that all basic mental representations are simple sensory experiences and all other mental representations derived from those. For example, Hume held that “simple ideas” are direct copies of simple sensory impressions and all other ideas obtained by some process of combination or association of these primary mental building blocks. Some ideas which are not basic on Hume's view are hard to reconstruct from simple impressions: for example, those of causation, time, object, and number.¹⁵ One might suspect that the PURE theory leads to comparable difficulties by way of the strong derivation thesis.

It is important to be clear on a crucial difference between strong derivation and Humean concept empiricism. The latter can be seen as a conjunction of two independent claims: first, that all mental states in some sense derive from experiences; second, that all experiences either are or are composed of simple sensory experiences. The PURE theory might reasonably be taken to suggest that all mental states derive from experiences (the first claim), but it is entirely compatible with the existence of non-sensory experiences (which would falsify the

second claim). The view suggested by the PURE theory is not Humean concept empiricism but one we could call *concept experientialism*.

Concept experientialism is arguably far more plausible than Humean concept empiricism. For a start, it seems that the class of basic, underived, phenomenally representable concepts is significantly larger than the Humean view allows. Crucially, it is plausible that the notions Hume had most difficulty reconstructing from sense impressions can be represented phenomenally. Take for example causation, time, and persistent objects. There is plausibly something it is like to notice a causal relation, be aware of the passage of time, or recognize a persistent object as such. Consequently, these notions do not have to be reconstructed from more basic ones on the concept experientialist view.

It is also worth stressing that the PURE theory and concept experientialism are compatible with a wide range of derivation mechanisms. For instance, a large number of entities might be picked out more or less descriptively through their effects or linguistic conventions (as Chalmers 2002b suggests). Other entities might be represented in virtue of inferential connections between certain states in the kind of way proponents of conceptual role semantics argue underpins all conceptual representation (e.g. Block 1986, Harman 1982). Perhaps mental representations can also get their contents deferentially without the intermediary of descriptive referential mechanisms (as Burge 1979 appears willing to argue). Given the potential ubiquity of intentional states which derive their contents through deference, composition, descriptive matching, or dispositions, it is plausible that only a small fraction of the contents we can entertain in thought are ever underivedly, phenomenally represented. One might think that concept experientialism preserves the germ of truth in Hume's first principle.

4.6. Unconscious mental states

An important objection to the PURE theory stems from unconscious mental states. It might seem obvious if not trivially true that unconscious states are not phenomenal. It might also seem probable that some unconscious states are underived intentional states. If both claims are true, it follows that some unconscious states are counterexamples to the derivation thesis and the PURE theory.

We need to be careful regarding how we understand the term “unconscious”. If it means the same as “non-phenomenal”, then it is trivial that unconscious states are not phenomenal. But to say that it is probable that there are unconscious underived intentional states would be to say simply that it is probable that there are non-phenomenal underived intentional states. The objection would boil down to a mere assertion that the derivation thesis is probably false; it would bring in no extra information that could overturn the evidence on the table for derivation. However, the objection does seem to throw new light on the derivation thesis given another interpretation of the term “unconscious” according to which it means *non-introspectible*. Since the non-introspectible states revealed by science are potentially quite different from ordinary introspectible states (on which I have focused throughout), one might reasonably suspect that they make exception to the derivation thesis, or to the premises I relied on in my argument for this thesis.

Putative examples of non-introspectible mental states fall into two broad categories I will label *subconscious attitudes* and *unconscious representational states*. Subconscious attitudes are states of the sort Freud postulated. They are attitudes like belief and desire which are not introspectible, or at least not in the normal way. Unconscious representational states are not as readily comparable to normal, introspectible states. Prime examples of unconscious representational states include the states constituted by early representations occurring along visual

pathways, dorsal stream visual representations, and “implicit” representations of grammatical rules. These three kinds of representation play important roles in current psychological theories, and they are commonly taken to constitute mental states even if they are entirely non-introspectible. One might argue that subconscious attitudes and unconscious representational states are counterexamples to the PURE theory because they can be both underived and non-phenomenal (which would contradict the derivation thesis).

In §3.2 I argued that attitudes are derived because they essentially involve cognitive dispositions. The same observation applies to subconscious attitudes, which are even more tightly connected with cognitive dispositions than introspectible attitudes. As a result, subconscious attitudes do not seem to pose a special difficulty at this stage. On the other hand, unconscious representational states could reasonably be taken to pose a fresh challenge: it might seem obvious that they are not phenomenal states, and it might seem plausible that they are underived intentional states because they appear to be neither attitudes nor composite states, hence not covered by the argument I gave in §3.2.

There are three complementary responses to this argument from unconscious representational states. The first is that it is not obvious that these states are intentional. To count as intentional states, the states constituted by unconscious representations would have to be mental states. One could challenge the common assumption that they are. Why should we not say simply that they are (non-mental) representational states causally involved in the generation of mental states? We do not yet have a well established reductive theory of intentionality. Lacking such a theory, we have to determine whether a state is intentional by comparing it with the paradigmatic intentional states. The paradigmatic intentional states are attitudes and experiences. There seems to be room for one to argue that the similarity between these and the bare unconscious information-carrying states postulated by cognitive scientists is too tenuous

to warrant calling the latter intentional states.

Another possible response to the objection from unconscious representational states is that, if they are intentional, then they are derived (hence compatible with derivation). Imagine the following experiment: we extract the early visual system of a mouse, including its eyes, and keep it functioning in isolation from the rest of the mouse. Would it seem appropriate to describe this isolated visual system as having intentional states? Probably not: it would seem farther from having intentional states than today's digital cameras. This suggests that it is at best when information-carrying states are suitably integrated within larger perceptual and cognitive systems that they deserve to be described as intentional states. Put differently, it seems that unconscious representational states could count as intentional only in virtue of being related to distinct intentional states in certain ways. This would mean that they are derived if they are intentional.

A third response is that, even if we grant that unconscious representational states are underived intentional states, there is no evidence that they are not phenomenal independently of *introspectible* counterexamples to the PURE theory. To begin with, it is not *a priori* that non-introspectible states are not phenomenal. For example, it seems perfectly conceivable that some animals have experiences but cannot introspect them for lack of introspective capabilities. This seems to be not only conceivable but actually true. It is doubtful that mice, for instance, can introspect their experiences. The common supposition that all phenomenal states are introspectible is also not supported by direct observation. Of course, all directly observable phenomenal states are introspectible. But this follows simply from the fact that we have no other way of observing mental states than through introspection. Normally, it is reasonable to conclude that all Fs are Gs from the fact that all observable Fs are Gs. But it would not be reasonable to infer that all Fs are observable from the fact that all observable Fs are observable. I

venture to say that a similar fallacy is behind much of the seeming plausibility of the claim that all experiences are introspectible.¹⁶

If the claim that all phenomenal states are introspectible is not supported by a priori reasoning or direct observation, it can only be supported by theoretical considerations: what we should say about non-introspectible states is simply what our best (most empirically adequate, simplest, etc) theory of consciousness implies about them. So to make a case against the PURE theory based on the claim that non-introspectible states are not phenomenal, one would first have to defend a competing theory which shows that non-introspectible states are not phenomenal. And one would have to do this based on observations of how introspectible phenomenal and non-phenomenal states relate to physical states. But I have already argued that the PURE theory offers the best account of consciousness as far as introspectible states go.

It is initially hard to believe that our brains might turn out to be replete with experiences we cannot introspect. But it is hard to believe that consciousness arises from the brain at all; non-introspectible representations do not seem much worse than other bits of the gelatinous structure of the brain *qua* candidate carriers of consciousness. In any case, my point is not that there actually are non-introspectible experiences. On the contrary, I recommend agnosticism on the matter. I believe that we ought to suspend judgment as to the existence of non-introspectible experiences until we have a well confirmed theory that allows us to settle the question. If the PURE theory is correct, we need a theory of underived intentionality before we can move forward on this front. The theory of intentionality we end up settling on might have the consequence that non-introspectible representational states are no more intentional or conscious than thermostats; alternatively, it might imply that each of us harbors far more phenomenal states than we ever suspected. The PURE theory by itself is neutral on this score.

5. The place of consciousness in nature

I would like to conclude by clarifying the import of the present view for the project of naturalizing consciousness.

The PURE theory bears primarily on the place of consciousness in the mind, not on its place in nature. It entails that consciousness is reducible if and only if underived intentionality is reducible, but it does not dictate whether both are reducible or neither is. It might nonetheless seem to have detrimental consequences for naturalization. First, it suggests the strong derivation thesis, which might seem to make consciousness metaphysically or explanatorily prior to intentionality. Second, it implies that phenomenal intentionality is altogether different from the (derived) intentionality of propositional attitudes. It might for both reasons seem to threaten the hope that representationalism would help the naturalization project by assimilating experiences to states we already know how to explain, namely, attitudes. Is not the upshot of the theory simply that consciousness is the sort of intentionality we cannot reduce?

The implications of the PURE theory for naturalization are not as bleak as the preceding might suggest. First, the view does not make consciousness prior to intentionality generally: nothing is prior to itself, so consciousness is at least not prior to underived intentionality on the PURE view. Regarding the sharp distinction it draws between attitudes and experiences, it is true that the theory rules out some accounts of intentionality as accounts of consciousness. For example, interpretationism and conceptual role semantics, which were designed with attitudes in mind, are nonstarters as accounts of *underived* intentionality. They are not candidate explanations of consciousness on the PURE view. But some theories of content could conceivably explain underived intentionality, for example, causal and/or informational theories. More importantly, any complete theory of intentionality should explain underived intentionality. This means that any complete theory of intentionality should by itself explain conscious-

ness if the PURE view is correct. On the PURE theory, there are no special ingredients which turn intentional states into experiences: all there is to consciousness is the most basic kind of intentionality.

Having said this, there is at least one way the strong derivation thesis suggested by the PURE theory could be used as part of an argument against reductionism: someone who holds that consciousness is irreducible could use this thesis to argue that all forms of intentionality are irreducible, thereby consolidating the anti-reductionist position. The PURE theory does not purport to settle the debate between reductionists and anti-reductionists, but it could in principle provide each camp with novel premises to draw on.

References

- Armstrong, David M. (1973). *Truth, Belief, and Knowledge*. Cambridge University Press.
- Audi, Robert (1972). "The concept of 'believing'." *The Personalist* 53:43-52.
- Baker, Lynne Rudder (1995). *Explaining Attitudes*. Cambridge University Press.
- Block, Ned (1978). "Troubles with functionalism." *Minnesota Studies in the Philosophy of Science* 9:261-325.
- Block, Ned (1986). "Advertisement for a semantics for psychology." *Midwest Studies in Philosophy* 10:615-78.
- Block, Ned (1993). "Holism, hyper-analyticity and hyper-compositionality." *Mind and Language* 8(1):1-26.
- Block, Ned (1995). "On a confusion about a function of consciousness." *Behavioral and Brain Sciences* 18:227-47.
- Burge, Tyler (1979). "Individualism and the mental." *Midwest Studies in Philosophy* 4:73-122.
- Byrne, Alex (2001). "Intentionalism defended." *Philosophical Review* 110 (2) 199-240.
- Byrne, Alex (2003). "Consciousness and nonconceptual content." *Philosophical Studies* 113(3):261-274.
- Chalmers, David J. (2002a). "The components of content." In David J. Chalmers, *Philosophy of Mind: Classical and Contemporary Readings*. Oxford University Press, 2002.
- Chalmers, David J. (2002b). "On sense and intension." *Philosophical Perspectives* 16:135-82.
- Chalmers, David J. (2004). "The representational character of experience." In Brian Leiter, *The Future for Philosophy*. Oxford University Press, 2004.
- Chalmers, David J. (2006). "Perception and the fall from Eden." In Tamar S. Gendler & John Hawthorne, *Perceptual Experience*. Oxford University Press, 2006.
- Cooney, Jeffrey W. & Gazzaniga, Michael S. (2003). "Neurological disorders and the structure of human consciousness." *Trends in Cognitive Science* 7(4):161-165.
- Crane, Tim (2003). "The intentional structure of consciousness." In Quentin Smith & Aleksandar Jokic (eds.), *Consciousness: New Philosophical Perspectives*. Oxford University Press.
- Davidson, Donald (1984). *Inquiries Into Truth And Interpretation*. Oxford University Press.
- de Sousa, Ronald (2007). "Emotion", *Stanford Encyclopedia of Philosophy*.
- Dennett, Daniel C. (1987). *The Intentional Stance*. MIT Press.
- Dretske, Fred (1995). *Naturalizing the Mind*. MIT Press.

- Dretske, Fred (2003). "Experience as representation." Philosophical Issues 13:67-82.
- Farah, Martha J. (1990). Visual Agnosia: Disorders of Object Recognition and What They Tell Us About Normal Vision. MIT Press.
- Fodor, Jerry A. (1990). A Theory of Content and Other Essays. MIT Press.
- Fodor, Jerry A. & LePore, Ernest (1992). Holism: A Shopper's Guide. Blackwell.
- Graham, George; Horgan, Terence E. & Tienson, John L. (2007). "Consciousness and intentionality." In Max Velmans & Susan Schneider, The Blackwell Companion to Consciousness. Blackwell, 2007.
- Harman, Gilbert (1973). Thought. Princeton University Press.
- Harman, Gilbert (1982). "Conceptual role semantics." Notre Dame Journal of Formal Logic 28(April):242-56
- Heck, Richard G. (2000). "Nonconceptual content and the space of reasons." Philosophical Review 109(4):483-523.
- Jackson, Frank (2003). "Narrow content and representation--or twin earth revisited." Proceedings and Addresses of the American Philosophical Association 77(2):55-70.
- Jackson, Frank (2004). "Representation and experience." In Hugh Clapin, Representation in Mind. Elsevier, 2004.
- Kriegel, Uriah (2002). "PANIC theory and the prospects for a representational theory of phenomenal consciousness." Philosophical Psychology 15(1):55-64.
- Kriegel, Uriah (2003). "Is intentionality dependent upon consciousness?" Philosophical Studies 116(3):271-307.
- Levin, Daniel T.; Simons, Daniel J.; Angelone, Bonnie L. & Chabris, Christopher (2002). "Memory for centrally attended changing objects in an incidental real-world change detection paradigm." British Journal Of Psychology 93:289-302.
- Lewis, David (1974). "Radical interpretation." Synthese 27(July-August):331-344
- Loar, Brian (1987). "Social content and psychological content." In Robert H. Grimm & D. D. Merrill, Contents of Thought. University of Arizona Press, 1987.
- Lycan, William G. (1982). "Toward a homuncular theory of believing." Cognition and Brain Theory 4:139-59.
- Lycan, William G. (1996). Consciousness and Experience. MIT Press.
- McGinn, Colin (1988). "Consciousness and content." Proceedings of the British Academy 74:219-39.
- Millikan, Ruth G. (1984). Language, Thought and Other Biological Categories. MIT Press.
- Nordby, Knut (1990). "Vision in a complete achromat: A personal account." In R. F. Hess, L. T. Sharpe & K. Nordby, Night Vision: Basic, Clinical and Applied Aspects. Cambridge University Press, 1990.
- Pettit, Philip (1993). The Common Mind. Oxford University Press.
- Prinz, Jesse J. (2002). Furnishing the Mind: Concepts and Their Perceptual Basis. MIT Press.
- Putnam, Hilary (1975). "The meaning of 'meaning'." Minnesota Studies in the Philosophy of Science 7:131-193.
- Rosenthal, David M. (1986). "Two concepts of consciousness." Philosophical Studies 49(May):329-59.
- Seager, William E. (1999). Theories of Consciousness: An Introduction and Assessment. Routledge.
- Seager, William E. (2003). "Tye on consciousness: Time to panic?" Philosophical Studies 113(3):237-247.
- Seager, William E. & Bourget, David (2007). "Representationalism about consciousness." In Max Velmans & Susan Schneider, The Blackwell Companion to Consciousness. Blackwell, 2007.
- Searle, John R. (1980). "Minds, brains and programs." Behavioral and Brain Sciences 3:417-57.
- Searle, John R. (1983). Intentionality: An Essay in the Philosophy of Mind. Cambridge University Press.
- Searle, John R. (1990). "Consciousness, explanatory inversion and cognitive science." Behavioral and Brain Sciences 13:585-642.
- Searle, John R. (1992). The Rediscovery of the Mind. MIT Press.
- Segal, Gabriel (2000). A Slim Book About Narrow Content. MIT Press.
- Stalnaker, Robert (1984). "Inquiry." Cambridge University Press.
- Strawson, Galen (1994). Mental Reality. MIT Press.

- Tye, Michael (1995). Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind. MIT Press.
- Tye, Michael (2000). Consciousness, Color, and Content. MIT Press.
- Venneri, Annalena & Shanks, Michael F. (2004). "Belief and awareness: Reflections on a case of persistent anosognosia." Neuropsychologia 42(2):230-238.

* David Chalmers has helped me a great deal with this article, its predecessors, and much related work. This project might well never have panned out without his many suggestions and constructive criticisms. Thanks also to the following for their helpful comments on ancestors of this paper: Tim Bayne, Ben Blumson, Alex Byrne, Magdalena Balcerak Jackson, Frank Jackson, Uriah Kriegel, Heather Logue, Fiona Macpherson, Farid Masrour, Adam Pautz, Nico Silins, Declan Smithies, Paul Snowdon, Daniel Stoljar, Amie Thomasson. Early discussions with Michael Lachelt, Alex Lamey, Tom Rand, Brad Richards and Joel Walmsley have also helped. Finally, thanks to William Seager for patiently enduring my very first attempts to articulate the main points of this paper and pointing me in fruitful directions.

¹ Dretske's proposal is also that experiences are abstract, nonconceptual and poised states, but he explains these conditions differently. According to him, a nonconceptual state is one that has its “indicator function” in virtue of natural selection (as opposed to some other kind of selection). Poised states are states which are prone to impact on belief (unlike Tye, Dretske does not say where they arise).

² This is also suggested by cases of anosognosia like that described in Venneri & Shanks (2004). For more detailed criticisms of the PANIC theory, see Byrne (2003), Kriegel (2002), Seager (1999, 2003), and Seager & Bourget (2007).

³ Alternatively, we could say that a representation is derived just in case it is such as to always obtain at least in part *in virtue of* distinct intentional states, or relations to such states. In this paper I use “obtains in virtue of” and “is realized by” interchangeably.

⁴ One might think that all the examples of derived states in this section are derived through composition. This is compatible with everything I say here, as one state can derive from other states in more than one way.

⁵ Matched states don't have to be narrow. Likewise, matching states don't have to be wide. I only use the narrow / wide distinction to illustrate the matching / matched distinction. Note also that I speak of matching and matched states, as opposed to matching and matched *contents* of a single state. There are states with both matched and matching contents (we can slice and recombine in-

tentional states as much as we want, as long as the resulting states involve intentional directedness among other things). However, the simplest states are those with a single content. I focus on those because states with distinct matching and matched contents would automatically be derived on my definition (they would be composite states), so they are of little interest here.

- ⁶ A possible objection here is that a phenomenal state *s* entails any state expressible as “being in *s* or ___”, so not all phenomenal states can occur independently of all other possible intentional states. I am not convinced that we should count disjunctive states as intentional, but I can grant this for the sake of argument. We need only qualify the previous claim: all phenomenal states can occur independently of all other phenomenal states they do not realize. This is enough to show that all phenomenal states are underived, because realization is anti-symmetric.
- ⁷ Heck (2000) does not gloss the state view as being that experiences are necessarily realized by concepts, but only as the view that experiences *depend* on concepts. I make it stronger out of charity to potential objectors to the PURE theory. Otherwise, another reply one could make to the objection that the PURE theory is incompatible with conceptualism is that dependency does not entail realization.
- ⁸ All the examples of presentations Brentano gives are conscious states: “By presentation I do not mean that which is presented, but rather the act of presentation. Thus, hearing a sound, seeing a colored object, feeling warmth or cold, as well as similar states of imagination are examples of what I mean by this term.” (*Psychology from an Empirical Standpoint*, L. McAlister, Ed. Excerpt from Chalmers 2002, *Philosophy of Mind: Classical and Contemporary Readings*, Oxford University Press, p. 479.)
- ⁹ “We shall say that we have *acquaintance* with anything of which we are directly aware, without the intermediary of any process of inference or any knowledge of truths. Thus in the presence of my table I am acquainted with the sense-data that make up the appearance of my table—its colour, shape, hardness, smoothness, etc.; all these are things of which I am immediately conscious when I am seeing and touching my table.” (*Ten Problems of Philosophy*, Oxford University Press, 1973, p. 25.)

- ¹⁰ It has mental states (or would if it existed) because intentional states are a species of mental state on the definition I gave in introduction.
- ¹¹ Thanks to David Chalmers and Heather Logue for emphasizing this objection.
- ¹² This objection was pointed out to me by Ben Blumson, David Chalmers, Farid Masrour, and Amie Thomasson.
- ¹³ Searle (1983) might be taken to be doing this. He claims that *perceptual content* is determined both by phenomenology and background beliefs and abilities. But this point does not bear on experiences *per se*. For if perceptual content is something more than experience (as the point implies), then its being derived does not imply that experience is derived. Searle at best gives us reason to distinguish experiential content from what he describes as perceptual content.
- ¹⁴ This suggests another objection to originality: it might seem that total experiences like *o* are derived through composition from their “components”. In response to this, the relation between total experiences and their so-called components is arguably like that between an experience of a red square and the experiences of redness and squareness: total experiences are determinates of their components (their determinables). Determinables do not realize their determinates. For example, the property of being moderately tall does not realize the property of being exactly six foot tall. So the components of total experiences do not realize them. Also, if one insists that determinables at least sometimes realize their determinates, we can amend the definition of “derived” to allow for this, since realization by determinables would seem entirely different from other kinds of realization.
- ¹⁵ See Prinz (2002) for a recent introduction to these classic problems, and potential solutions.
- ¹⁶ Rosenthal (1986) makes a similar point, albeit in a different context.